What is claimed is:

SUB Cl7

1. A computer-assisted method for evaluating a cluster assignment for an observation, comprising the activities of:

for each of a plurality of observations, obtaining a data set containing no more than one proxy value for each of a plurality of variables, each variable having a plurality of possible values, the data set also containing a cluster assignment for the observation, the cluster assignment identifying one cluster from a plurality of clusters;

for each observation from the plurality of observations, calculating a percent of proxy values for the plurality of variables that equals a mode of that observation's corresponding cluster's proxy values for the corresponding variables; and

outputting the percent for each observation.

2. A computer-assisted method for evaluating a cluster assignment for an observation, comprising the activities of:

for each of a plurality of observations, obtaining a data set containing no more than one proxy value for each of a plurality of variables, each variable having a plurality of possible values, the data set also containing a cluster assignment for the observation;

for each observation from the plurality of observations, estimating a purposeful probability that a particular possible value from the plurality of possible values for a particular variable will be purposefully provided by observations assigned to a particular cluster from a plurality of clusters; and outputting each purposeful probability.

3. The method of claim 89, further comprising the activities of:

for each observation from the plurality of observations in each cluster from the plurality of clusters, calculating a serendipity probability for each possible value, the serendipity probability a measure of a probability that an observation in a particular cluster will be randomly associated with any one of the plurality of possible values for a particular variable;

for each observation from the plurality of observations, calculating a ratio of the purposeful probability to the serendipity probability;

for each observation from the plurality of observations, calculating a logarithm of the ratio to obtain composition analysis score; and

outputting the composition analysis scores for each observation in each cluster.

4. The method of claim 89, further comprising the activities of:

for each observation from the plurality of observations, assuming that before the observation can be made, the observation has an equal probability of being in any identified cluster from the plurality of clusters;

for each observation from the plurality of observations, assuming that the purposeful probabilities are true;

for each observation from the plurality of observations, using Bayes' Theorem to calculate a Bayes probability that a particular observation can be in each cluster conditional upon the observation's proxy value to each variable;

outputting the Bayes probability that each observation can be in each cluster.

5. A computer-readable medium containing instructions for activities comprising:

for each of a plurality of observations, obtaining a data set containing no more than one proxy value for each of a plurality of variables, each variable having a plurality of possible values, the data set also containing a cluster assignment for

the observation, the cluster assignment identifying one cluster from a plurality of clusters;

for each observation from the plurality of observations, calculating a percent of proxy values for the plurality of variables that equals a mode of that observation's corresponding cluster's proxy values for the corresponding variables; and

outputting the percent for each observation.

6. An apparatus for evaluating a cluster assignment for an observation, comprising:

for each of a plurality of observations, means for obtaining a data set containing no more than one proxy value for each of a plurality of variables, each variable having a plurality of possible values, the data set also containing a cluster assignment for the observation, the cluster assignment identifying one cluster from a plurality of clusters;

for each observation from the plurality of observations, means for calculating a percent of proxy values for the plurality of variables that equals a mode of that observation's corresponding cluster's proxy values for the corresponding variables; and

means for outputting the percent for each observation.

7. A computer-readable medium containing instructions for activities comprising:

for each of a plurality of observations, obtaining a data set containing no more than one proxy value for each of a plurality of variables, each variable having a plurality of possible values, the data set also containing a cluster assignment for the observation;

for each observation from the plurality of observations, estimating a purposeful probability that a particular possible value from the plurality of possible values for a particular variable will be purposefully provided by observations assigned to a particular cluster from a plurality of clusters; and

outputting each purposeful probability.

8. An apparatus for evaluating a cluster assignment for an observation, comprising:

for each of a plurality of observations, means for obtaining a data set containing no more than one proxy value for each of a plurality of variables, each variable having a plurality of possible values, the data set also containing a cluster assignment for the observation;

for each observation from the plurality of observations, means for estimating a purposeful probability that a particular possible value from the plurality of possible values for a particular variable will be purposefully provided by observations assigned to a particular cluster from a plurality of clusters; and

means for outputting each purposeful probability.